CHAPTER TWO

Sonata Form as a Whole

Foundational Considerations

Sonata form is the most important large structure of individual movements from the "common-practice" tonal era. It sets forth and resolves its musical discourse within a large-scale binary format. The term "sonata form" was almost surely unknown to Haydn, Mozart, early Beethoven, and their contemporaries: it seems to have surfaced only in the 1820s and 1830s. In the late-eighteenth and early-nineteenth century this structure would have been grasped primarily as the customary design of first movements within sonatas, chamber music, and symphonies, although it was by no means confined only to first movements (nor only to rapid-tempo movements). The varying descriptions from contemporary theorists were more convoluted. There the form was variously described as: "the first allegro of the symphony [or sonata]" disposed in "two sections" [*zwey Theile*] and three "main periods" [*Hauptperioden*] (Koch 1793); within "larger pieces of music" a "well-conducted melody [!] . . . divided into two parts, either connected, or separated in the middle by a repeat sign" (Galeazzi 1796); "an elaborate movement [or] a long movement . . . generally divided into *two sections*" (Kollmann 1799); "grand binary form" [*grande coupe binaire*] (Reicha 1826); and so on.[1] Still, "sonata form" (*Sonatenform*) seems to have been a familiar term by the mid-1820s, at least in A. B. Marx's *Berliner allgemeine musikalische Zeitung* circle, where it referred both to the multimovement cycle as a whole and, occasionally, to the form of an individual movement.[2] It was only in 1838 and 1845, though, in technical

1. Heinrich Christoph Koch, *Versuch einer Anleitung zur Composition* (Leipzig: Adam Friedrich Böhme, 1793; rpt., Hildesheim: Georg Olms, 1969), pp. 304–5 (from section 101), trans. Nancy Kovaleff Baker in Koch, *Introductory Essay on Composition: The Mechanical Rules of Melody, Sections 3 and 4* (New Haven, Conn.: Yale University Press, 1983), p. 199; Francesco Galeazzi, *Elementi teorico-pratici di musica*, vol. 2 (Rome: Puccinelli, 1796), the relevant extracts of which were excerpted and translated in Bathia Churgin, "Francesco Galeazzi's Description (1796) of Sonata Form," *Journal of the American Musicological Society* 21 (1968), 181–99 (above quotations from pp. 189–90); A. F. C. Kollmann, *An Essay on Practical Musical Composition* (London, 1799; rpt. New York: Da Capo Press, 1973), p. 4 [ch. 1, section 10]; Anton Reicha, *Traité de haute composition musicale* (Paris, 1826), discussed, e.g., in Ian Bent and William Drabkin, *Analysis* (New York: Norton, 1987), pp. 18–20, and especially Peter A. Hoyt, "The Concept of *développement* in the Early Nineteenth Century," in *Music Theory in the Age of Romanticism*, ed. Ian Bent (Cambridge: Cambridge University Press, 1996), pp. 141–62.

2. In the journal's first year of publication (1824) the term 'sonata form" appeared in both senses. The first, apparently initially the more common, was a description of the entire multimovement cycle (used by Marx,

discussions of the form's particulars, that Marx put the stamp of approval on the term "Sonatenform" with regard to the individual-movement structure.[3] Throughout this book we use that term as a familiar quick-reference, even as we realize that that designation was not current in the eighteenth century.

Sonata form is neither a set of "textbook" rules nor a fixed scheme. Rather, it is a constellation of normative and optional procedures that are flexible in their realization—a field of enabling and constraining guidelines applied in the production and interpretation of a familiar compositional shape. Existing at any given moment, synchronically, as a mappable constellation (although displaying variants from one location to another, from one composer to another), the genre was subjected to ongoing diachronic transformation in history, changing via incremental nuances from decade to decade. Haydn's conception of what was customary within sonata form in 1770 differed somewhat from Beethoven's conception in 1805. However such models might be said to have differed, they also shared certain crucial, genre-defining features that make them all recognizable as sonata form. Here we are dealing primarily with the model that crystallized during the second half of the eighteenth century and that reached a peak in the mature works of Haydn and Mozart and the early works of Beethoven.

What we now call sonata form was developed as a response to aspects of the world view of the Enlightenment and the concomitantly emerging modernism. Considered generally, it could be understood as an abstract metaphor for disciplined, balanced action in the world, a generalized action involving differing types of idealized mid- and late-eighteenth-century personalities. (Its potential for "extramusical" connotations and analogues is discussed in the final section of chapter 11, "Narrative Implications: The Sonata as Metaphor for Human Action.") Sonata form emphasized short-range topical flexibility, grace, and forward-driving dynamism combined—in both the short and long range—with balance, symmetry, closure, and the rational resolution of tensions. By the mid-eighteenth century it had become obligatory for the first movement of a standard multimovement instrumental work; it had also become a common, if optional, choice for the slow movement and the finale. Slow movements and finales sometimes also displayed different adaptations of the form. Although the guidelines in most of this book were written predominantly with first and last movements and single-movement overtures in mind (all energetic "Allegro movements"), they are also applicable, occasionally with some modifications, to slow movements.

From the compositional point of view sonata form was an ordered system of generically available options permitting the spanning of ever larger expanses of time. A sonata-form project was a feat of engineering, like the construction of a bridge "thrown out" into space. In the eighteenth-century style this temporal span was to be built from rather simple materials: trim, elementary musical modules whose brevity and small-scale balances seemed best suited to short-winded compositions. In the hands of most composers, constructing a sonata-form movement was a task of *modular assembly*: the forging of a succession of short, section-specific

Heinrich Joseph Birnbach, and others), a usage that persisted throughout much of the nineteenth century, especially in German-speaking regions. The other use of "Sonatenform" referred to the structure of an individual movement. It first appeared in a casual, unexplained way—as if it were already a common label—in Marx's 1824 essay on the E-minor second movement (Prestissimo) of Beethoven's Piano Sonata in E, op. 109 ("Es bildet mit dem letzen Satze die eigentliche Sonate und ist auch in der Sonaten-Form hingeworfen," *BamZ*, I, 1824, 37b) and in Carl Loewe's discussion of the first movement of Beethoven's Cello Sonata, op. 102 no.1 ("Hart und rauh, im männlichen Zorne, beginnt ein kurzes Allegro (A-moll) in der Sonatenform," *BamZ*, 1824, 410b). See the discussion of terminology and quotation of sources in the entry by Hans-Joachim Hinrichsen, "Sonatenform, Sonatenhauptsatzform" [1996], in Hans Heinrich Eggebrecht, ed., *Handwörterbuch der musikalischen Terminologie* (Stuttgart: Steiner, n.d.), pp. 1–7.

3. A. B. Marx, *Die Lehre von der musikalischen Komposition, praktisch-theoretisch*, vols. 2 and 3, 1st eds. (Leipzig, 1838 and 1845), 2:482, 497; 3:195; cited in Hinrichsen, "Sonatenform, Sonatenhauptsatzform" [1996], pp. 6–7. See also Marx, *Musical Form in the Age of Beethoven: Selected Writings on Theory and Method*, ed. and trans. Scott Burnham (Cambridge: Cambridge University Press, 1997).

musical units (spaces of action) linked together into an ongoing linear chain—pressing down and connecting one appropriately stylized musical tile after another.[4] One of the challenges facing the mid- and late-eighteenth-century composer was to use a seemingly unassuming, *galant* language, grounded in structural punctuation and periodicity, to produce ever more spectacular spans for occasions of enhanced dignity, prestige, or social importance. Ever-larger, thematically differentiated binary structures (sonata forms, often with built-in repetitions of individual sections), eventual accretions to the structure (slow introductions and longer codas), and multimovement conventions all had their roles to play in this process of generic enlargement. And ultimately they led to the grandly monumental, personalized structures of Haydn, Mozart, and Beethoven.

The most typical sonata forms (what we call Type 3 sonatas) articulate an overall rounded binary structure. The two parts of this larger structure are, in modern terminology: (1) the exposition and (2) the development and recapitulation. As will be elaborated at the end of this chapter, both parts may be marked for repeat, or the composer may eliminate the repeat of part 2 or, under some circumstances, both repeats. Notwithstanding its binary origins, the normative, Type 3 sonata consists of three musical action-spaces (again, the exposition, development, and recapitulation), laid out in a large A||BA' format. Hence the common observation that the form consists of an originally binary structure often arrayed in a ternary plan. Each of the three spaces is usually subjected to thematic and textural differentiation. Each is marked by several successive themes and textures, all of which are normally recognizable as generically appropriate for their specified location. These three spaces can be viewed as expansions of the three phases of the continuous rounded binary form (the rounded binary structure in which the first part ends in a secondary key). We shall take up these spaces individually. (In figure 2.1 we have provided two diagrams of Sonata Theory's conception of the most common type of sonata form: 2.1a refers to the exposition; 2.1b to the whole sonata-form movement.)

Exposition

As with all of the action-spaces the exposition is assigned a double-task, one harmonic and the other thematic-textural ("rhetorical"). Its *harmonic* task is to propose the initial tonic and then, following any number of normative (and dramatized) textural paths, to move to and cadence in a secondary key. In major-mode sonatas—the most common in the eighteenth century—this was the key of the dominant (which may be indicated as V_T, meaning "a V that is tonicized"), thereby generating tonal tension. In minor-mode sonatas this was usually the key of the major mediant (III), although a less-often-selected choice (second-level default) was the minor dominant (v). The differing psychological and structural world of minor-mode sonatas is dealt with in chapter 14. Here, for the most part, we shall focus on major-mode practice.

The exposition's *rhetorical* task, no less important, is to provide a referential arrangement or layout of specialized themes and textures against which the events of the two subsequent spaces—development and recapitulation—are to be measured and understood. We refer to this layout as *Rotation 1* or the *expositional rotation*.[5]

4. To be sure—and particularly in the hands of the master composers of the period—certain passages within individual sonata forms may from time to time give the impression of a broader continuity of internal ramification. This is especially the case with the startlingly original musical language of Haydn, who, even within a generally modular and "sectionalized" concept of formal practice, often favored passages of ongoing *Fortspinnung* (a moment-to-moment "spinning-out" of modular growth and elaboration). For brief characterizations of Haydn's often-"vitalistic" compositional style, see ch. 11, subsection "Recompositions, Reorderings, Interpolations" (especially n. 2 and the text to which it refers), and ch. 18, subsection "Haydn's Treatments of Type 4 Finales" (especially n. 49 and related text).

5. Sonata-form structures are centrally concerned with the formal principle that we call *rotational form* or the *rotational process*: two or more (varied) cyclings—*rotations*—through a modular pattern or succession laid down at the outset of the structure. Appendix 2 provides a broader introduction to this principle, which pervades the discussion of sonata form in this book.

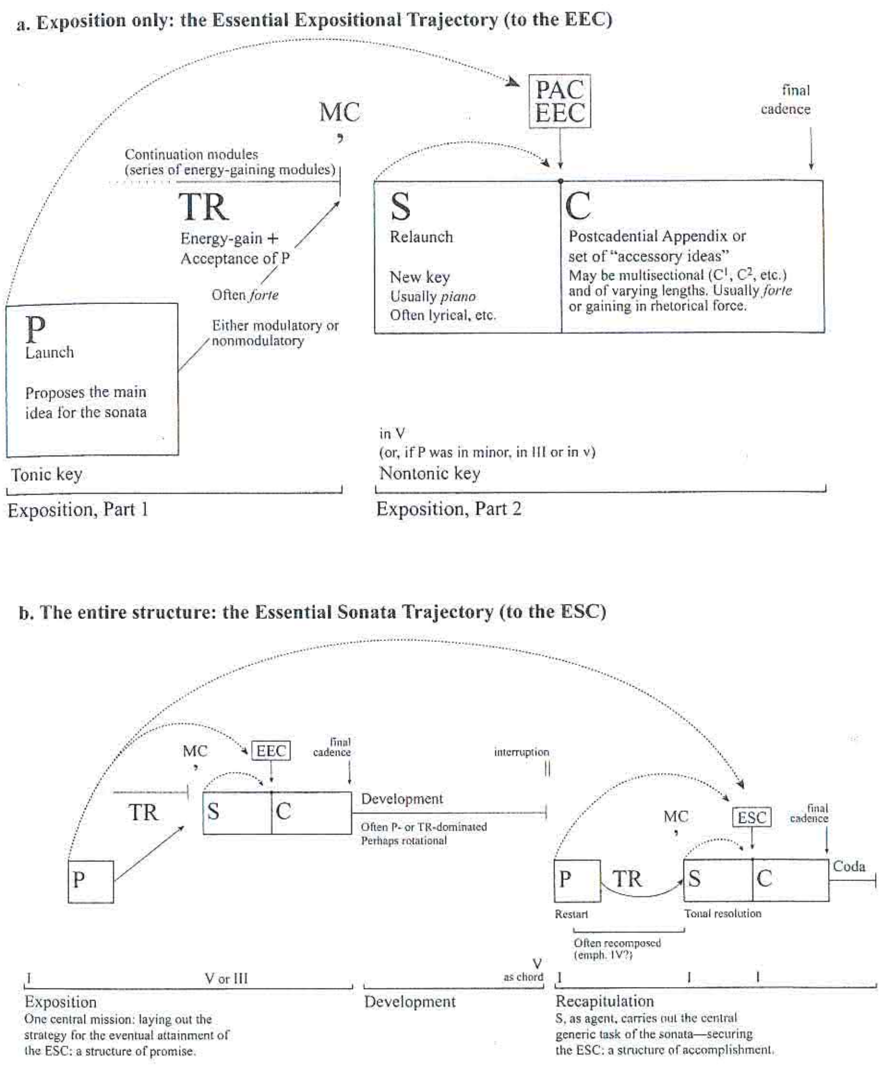

FIGURE 2.1 The Generic Layout of Sonata Form

Because the exposition's succession of events serves, especially in its second half, to predict the plan and purpose of the entire third space—the recapitulation, which finally resolves the work—its layout may be understood as articulating a *structure of promise* (indicating how it proposes that "things work out" in the recapitulatory rotation-to-come). Because the arrangement of rhetorical modules in Rotation 1 provides the ordered set of events that articulates the uniqueness and specific personality of that piece, it should be kept in mind when assessing all of the later events in the movement.

Within the expositional rotation the tonal and rhetorical tasks unfold simultaneously, intertwined with each other in mutually reinforcing ways. The exposition begins with a *primary theme* or *primary idea* (P) in the tonic that sets the emotional tone of the whole work. The most common layout for the remainder of the exposition continues with an energy-gaining zone of transition (TR) that leads to a mid-expositional break or medial caesura (MC). This is typically followed by the onset of a specialized, secondary-theme zone (S) in the new key. The generically essential tonal purpose of the exposition is to drive to and produce a secure perfect authentic cadence (PAC) in the new key (notated as V:PAC in major-mode sonatas, III:PAC or v:PAC in minor-mode ones). We refer to the first satisfactory PAC within the secondary key that goes on to differing material as the point of *essential expositional closure* (the EEC): this is one of the central concepts of Sonata Theory and one that is dealt with at length in other chapters.[6] Producing the EEC is the generically assigned task of the S-idea(s). The large dotted-line arrow in figure 2.1a suggests a broadly vectored trajectory from the start of the exposition to the EEC; the smaller dotted-line arrow below it suggests a subordinate trajectory from the beginning of S to its own point of PAC-closure at the EEC. In performing or listening to any sonata-form exposition one should sense the broad drive of these generic vectors. Whenever one hears the onset of S-space within any exposition, one should listen with an alert sense of anticipation for any subsequent PAC—how it might be approached, secured, delayed, thwarted, or deferred. One should experience any sonata form with a strongly "directed" preparatory set, pressing forward conceptually and anticipating genre-defining events-to-come.

Following the EEC one or more additional cadences (PACs) may follow within the *closing zone* or *closing space* (C). (Not all expositions contain C-modules; it is possible for the S-concluding EEC to be delayed until the end of the exposition, in which case there is no closing zone.) Whether or not C-modules are present, the final cadence of the exposition will generally be a perfect authentic cadence in the secondary key (again, V:PAC, III:PAC, or v:PAC). This final cadence might not occur directly at the double bar. Frequently the final cadence is followed by a C-module that prolongs the newly reinforced tonality by means of a pedal-point or some other device. Additionally, the final cadence is sometimes followed by a reactivation of V in preparation for a repeat of the entire exposition: if so, this reactivating passage is the *retransition* (RT).

Development

This action-space renders the established tonal tension more fluid and complex. While the exposition had split its tonal assertions into two broad blocks or contrasting planes (I and V in major-mode sonatas), the development typically initiates more active, restless, or frequent tonal shifts—a sense of comparative tonal instability. Here one gets the impression of a series of changing, coloristic moods or tonal adventures, often led (in major-mode works) through the submediant key, vi, or other minor-mode keys with shadowed, melancholy, or anxious connotations. Any authentic-cadence attainment in a non-tonic key is to be understood as an important developmental event—a cadential ratification of an attained tonal station. (A vi:PAC is especially common in major-mode sonatas.) Ultimately, the standard development culminates on an active dominant (V_A, meaning "a V that is an active chord, not a key"). At this point the dominant from the end of the (major-mode) exposition is usually recaptured, detonicized, and reactivated.

This last point needs underscoring. In the development the final cadence is usually a half-cadence in the tonic (I:HC), although a cadence in a related minor key, normally followed by a brief reactivation of V, is also a possibility. In addition, a I:HC is frequently followed by a prolongation of dominant harmony, a "dominant-lock" or "dominant preparation." The typical I:HC conclusion of the development—just before the onset of the recapitulation—brings us to a harmonic *interruption*. (This crucial interruption is a defining feature of the Schenkerian conception of sonata form.) The V_A at the end of the development is not resolved to the I that usually begins the recapitulation. Rather, the phrase—and the development section as a whole—is normally "interrupted" on V_A (notwithstanding any foregrounded or local, connective "fill" that might bridge the end of the development to the recapitulation), and the next cycle of events is newly launched with the opening of the recapitulation. True, this more fundamental interruption on the dominant may sometimes be masked on the foreground with an *apparent* V–I cadence (with the I triggering the recapitulation). But the more fundamental or background concept is that of harmonic interruption on V_A. (Those unfamiliar with the Schenkerian, linear-contrapuntal view of things might notice that this interruption divides the entire sonata form at the end of the development. This contrasts with the eighteenth-century "binary" division of sonata form at the end of the exposition.)

In terms of their *rhetorical* strategies, developments may or may not be fully or partially *rotational* (that is, guided in large part by the ordered thematic pattern established in the exposition). Developments often refer back to (or take up as topics) one or more of the ideas from the exposition, most commonly selected, as it happens, from Rotation 1's first half (P and TR). More often than not, the modules taken up and worked through in the development are presented in the order that they had originally appeared in the exposition (even though several expositional modules are normally left out entirely). Thus the modular succession encountered in the development—not only the expositional events referred to, but also the possibility of an episode or largely new theme—is never to be considered arbitrary. On the contrary, even within this more unpredictable, developmental texture the thematic choice and arrangement is of paramount importance and derives its significance through a comparison with what had happened in the exposition. The development is variable in length, although in the period 1760–90 one would normally expect it to occupy a smaller space than that established by the exposition. Longer, more elaborate developments in the 1780s, 1790s, and later decades are monumentalized statements that invite special attention.

Recapitulation

This action-space resolves the tonal tension originally generated in the exposition by rebeginning on the tonic (with the initial theme in the most common Sonata Types, 1, 3, 4, and 5) and usually by restating all of the non-tonic modules from part 2 of the exposition (S and C material) in the tonic key. For this reason—its largely referential retracing of the rhetorical materials laid out in the exposition (Rotation 1)—we also call the recapitulation the *recapitulatory rotation*. (Exceptions and reorderings of thematic material may be found in some sonatas.) Because of its function in bringing tonal closure to the entire form, we refer to the S/C complex in the recapitulation as the *tonal resolution*. Its shape and manner of unfolding had been established by the exposition's structure of promise. Correspondingly, we consider the recapitulation to articulate a *structure of accomplishment*. Minor-mode sonatas

6. For the moment, we might emphasize that the first *satisfactory* PAC in the new key is often but not always the *first* PAC in that key. A first PAC, for instance, might be followed by a thematic repetition of all or part of the S-idea that we have just heard—which would automatically defer the EEC to the next satisfactory PAC further ahead. Additionally, there are other ways of deferring the sense of a clear EEC (ch. 8). The clearest way of suggesting all of this in brief is to define the EEC as the first new-key PAC that *proceeds onward to differing or contrasting material*—or, of course, that closes the exposition itself, if there are no closing modules that follow that PAC.

that had sounded S and C in the major mediant (III) in the exposition have the additional option of sounding them in either the major or minor mode in the recapitulation.

The recapitulation's S, launching the tonal resolution following a recapitulatory MC, leads to the production of a satisfactory I:PAC that goes on to differing, non-S material. This is the moment of *essential structural closure* (the ESC), most often a point parallel to the exposition's EEC. The ESC represents the tonal goal of the entire sonata form, the tonal and cadential point toward which the trajectory of the whole movement had been driving: this is suggested by the longest dotted-line arrow in figure 2–1b. From the perspective of Sonata Theory, it is only here where the movement's tonic is fully called forth, stabilized as a reality as opposed to a mere potential. As in the exposition, C-material will follow, now in the tonic. The recapitulation's final cadence is generally a I:PAC (or, in minor, sometimes a i:PAC), although this too may be followed by a prolongation of tonic harmony or by a transition leading either back to a repeat (of the entire development and recapitulation) or forward into the coda.

A coda (outside of sonata space) may or may not follow the recapitulation. More information about codas, along with a discussion of the other optional or parageneric feature of some sonatas, the introduction, may be found in chapter 13.

Repetition Schemes

Within eighteenth-century sonatas and symphonies one may find both parts repeated (||: exposition :||: development–recapitulation :||). This is the most formal and earliest norm. Many late-century first movements, especially those after about 1760, repeat only the first part (the exposition), although in works prior to 1790 one need not be surprised to see the second part also repeated. After that date, repeating the second part is an uncommon gesture that invites analytical interpretation. It is also possible to find both parts unrepeated. This occurs in lighter works, in some midcentury symphonies (some Stamitz symphonies from the 1750s; some early Mozart symphonies; and so on) and in some slow movements (especially those in the format of the less expansive, Type 1 sonata, lacking a development). The nonrepeated exposition is also a generic feature of the overture or *sinfonia*. (In other words, expositional repeats will not appear in either operatic or concert overtures; this is also true of the overture's mid-nineteenth-century offspring, the symphonic poem).[7] In this aspect the lighter overture is to be distinguished generically from the more formal first movement of a sonata or grand symphony, which at least had available the common option of expositional repetition. Nonrepeated expositions within first movements do sometimes occur in more broadly scaled and ambitious works after 1780, but when they do—as in Mozart's Symphony No. 35 in D, K. 385, "Haffner,"[8] or in Beethoven's Violin Sonata in C Minor, op. 30 no. 2, his Piano Sonata in F Minor, op. 57, "Appassionata," and his String Quartet in F, op. 59 no. 1—they are exceptional and need to be considered as consciously expressive choices.[9]

One curious (and rare) possibility is that of literally writing out an expositional repeat,

7. Thus the rule. Exceptions are extremely rare and disconcertingly puzzling, such as the repeat of the exposition in young Mozart's Overture to *Apollo et Hyacinthus*, K. 38 (1767), labeled as the "Prologus/Intrada" to the opera. This piece is a Type 2 sonata (Chapter 17) whose first rotation (exposition) is provided with a repeat sign. Much later, the odd "expositional" (?) written-out and slightly varied repetition in Berlioz's Overture, *Le carnaval romain* is also curious, suggesting that the form of this unusual piece is more purely rotational (or perhaps instrumental-strophic with *fortissimo* refrain) than a sonata per se, although it is also manifestly in dialogue with certain sonata norms.

8. Other examples within Mozart's major works include the first movements of his Symphonies Nos. 31 in D, K. 297, "Paris," and 34 in C, K. 338, along with those of the Serenades in D, K. 320, "Posthorn," and in E-flat, K. 375. Such examples—perhaps related to earlier or existing concepts of repeat-convention options in overture-symphonies, in smaller-scale symphonies, or in some serenades—require individual attention. Within the larger symphony it may be that during the 1770s (though not, it seems, in the 1780s) Mozart was exploring the possibility of the omission of the expositional repeat as a lower-level default.

9. The solution of Beethoven's op. 59 no. 1/i, which initially suggests an expositional repeat only to abort it almost immediately in favor of development, is anticipated in the first movements of Mozart's Serenade in E-flat for Eight Winds, K. 375, and Haydn's Piano Sonata in D, Hob. XVI:51.

normally including variants the second time around. This occurred most famously in C. P. E. Bach's unusual set of six keyboard *Sonaten mit veränderten Reprisen*, H. 136–39, 126, 140 (W. 50/1–6, Sonatas with Varied Repetitions), composed in 1758–59 and published in Berlin the following year. In Haydn's works the procedure surfaces only (and wisely, in Tovey's view) in a few "purely lyric slow movements," such as the Adagios of the Quartet in C, op. 33 no. 3, "Bird," and the Symphony No. 102 in B-flat.[10] (Both slow movements are in F major; in the quartet the Adagio is the third movement; in the symphony it is the second.)

What are the purposes of large-scale repeats within sonata form? Central to the concept of the grand sonata or symphony is a system of schematic repeat-conventions, balances, symmetries and proportions that call attention to and help to define the genre. The emphatically architectural construction calls attention to the genre's ordered formality—and in the case of the grand symphony, also to its grandeur and public splendor. Repeats were an important feature of a sumptuous, high-prestige display of grand architecture, one to which large-scale repetitions were essential —especially that of the expositional repeat in the first movement. The stylized form thus celebrated the "Enlightenment" (or "modern") culture that makes such an impressive, moving, or powerful art possible. One of the structure's implications would have been that this culture had devised a rational, balanced means to shape and contain the fluid, raw, elemental power of music. By extension, the process probably also represented the controlling or harnessing of those impulsive, instinctive, libidinal, or "uncivilized" elements within ourselves. Control, balance, generic identification, and formal architectural splendor: these would appear to be the central reasons why literal repetition played such a prominent role in the style.

Consequently, repeat signs should not be taken for granted, passed over lightly in analysis, or omitted in performance. Repeat signs are never insignificant.[11] Block-repetitions are an integral component of the style, and composers can work with this defining convention in a variety of ways. When previously obligatory (or exceptionally strong first-level default) expositional repeats began gradually to disappear—especially in the early nineteenth century, with certain works of Beethoven (op. 30 no. 2, op. 57, op. 59 no. 1, and so on, and later with Mendelssohn, Schumann, and others)—the genre itself was undergoing a major rethinking.[12] The familiar, current views—Schenkerian and otherwise—that propose that some repeats are structurally insignificant while others are more important (because of the unfolding of certain structural tones or other significant events, perhaps under a first-ending sign) miss the larger point of repeat signs as generic identifiers.[13] Even when the structural-tone aspects might be convincing (but, perhaps paradoxically, only as local details), the gist of these claims seems to be based on later-nineteenth-century premises, which came to look on all unaltered repetition as an aesthetic error. Such a conviction also came to affect performance in the omission of repeats or in the insistence on an altered interpretation in the repeat. It may be, though, that saying the same thing twice was what the composer had in mind.

It is easy to object to our general argument here. One could strive to minimize the impor-

10. Tovey, "Sonata Forms," *Musical Articles*, p. 214: "Haydn saw that the only place for C. P. E. Bach's device was in purely lyric slow movements. Even there he never had the patience to plod and pose (as C. P. E. Bach did to the bitter end) through a repetition [recapitulation] of both parts. When his second part comes to recapitulate the second group it combines both versions."

11. For the quintessential statement of that which the present argument opposes, see Douglass M. Green, *Form in Tonal Music: An Introduction to Analysis*, 2nd ed. (New York: Holt, Rinehart, and Winston, 1979), p. 82: "HISTORICAL NOTE [*sic:*] Ordinarily the repetition of a part is of little significance in formal analysis."

12. Curiously, in 1826 Reicha suggested—in passing and without explanation (*Traité de haute composition musicale*, p. 300)—that finales may lack an explicit repeat: "When the first part is not repeated, as in overtures and finales . . ." ("Quand la première partie n'a pas de reprise, comme dans les ouvertures ou dans les finales . . ."). It may be that Reicha had sonata-rondos, Type 4 sonatas, in mind (ch. 18).

13. Cf., e.g., Jonathan Dunsby, "The Formal Repeat," *Journal of the Royal Musical Association* 112/2 (1987), 196–207.

tance of the usual repetition schemes by an appeal to history: deriving them step-by-step from the earlier binary forms, then asserting that the persistent lingering of the repeat conventions into the 1780–1820 period of the grand symphony was an outdated survival, vestigial, unnecessary to the perception of the genre. The larger question, though, is why the convention remained available into the later phases of 1780–1820 period and beyond (particularly after Beethoven's occasional removals of the expositional repeat had occurred). The expositional repeat must have persisted, however sporadically, because it was not merely vestigial. It continued to be genre-defining, a sign of special grandeur and formality—with an ear attuned also to the grand tradition and historical lineage that had led to the mid- and later-nineteenth-century sonata and symphony.

Of the two standard large-scale repeats, the second, longer one (development-recapitulation) was the one more vulnerable to suppression. This second repetition was increasingly reduced to the status of an easily discardable option in the 1780–1800 period.[14] In some cases, concerns of absolute length or a sense of redundancy in closing particularly dramatic sonatas twice might have overridden the genre-defining principle of long-range architectural repetition and balance, at least with regard to this development–recapitulation section. Perhaps the logic of the situation suggested that the obligatory repeat of section 1 alone (the expositional repeat) was to be viewed as sufficient as a genre definer.

However we decide this matter, we should note three things. First, the issue of notationally indicating a repeat of section 2 was still part of the historical concept of "grand binary" form (within a symphonic first movement) around 1800, even when that repeat was notationally elided. Its conceptual presence remained there, counterpointed against the given, simpler structure. It persisted as historical-generic memory, even when it was not made physically present on the acoustic surface of the music. Second, any retention of the second repeat toward the end of the eighteenth century should be regarded as expressively significant, especially since its strongest composers—Haydn and Mozart—were apparently coming to believe that repeat 2 was not as obligatory as that of repeat 1. When the repeat was called for, it must have been placed there for a reason, as in the slow movement and finale of Mozart's Symphony No. 41 in C, K. 551 ("Jupiter"), where formal processes and monumentalized grandeur are principal topics throughout the whole work. Third, given a nineteenth-century work lacking an indication of that second block-repetition, any reworked referencing back to this increasingly atavistic repeat 2 within a longer, discursive coda, as in the first movement of Beethoven's Symphony No. 3 in E-flat, op. 55, "Eroica," should be viewed as such, not as an innovative addition or accretion to a previously postulated, differing symphonic practice.

14. See, e.g, Michael Broyles, "Organic Form and the Binary Repeat," *The Musical Quarterly* 66 (1980), 339–60.

CHAPTER THREE

The Medial Caesura and the Two-Part Exposition

An exposition has both a tonal and a rhetorical function. Its basic tonal plot—moving from an initial tonic to a secondary key, then securing that new key with one or more cadences—constitutes the exposition's *tonal form.* This tonal form, generally the same in all sonatas, is worked out in different sonatas in individualized ways, according to localized rhetorical plots. Tonal form is to be distinguished from *rhetorical form*, which includes personalized factors of design and *ad hoc* expression: modular and textural layout, selection and arrangement of musical topics, varieties of structural punctuation, and so on. The compositional ordering of these processes produces a distinct, singular musical shape. This layout serves as the *referential rotation* (or *expositional rotation*) that also guides our understanding of the ordering of modular events in the subsequent action-spaces of the sonata—development, recapitulation, and coda. An exposition may be disposed in either of two rhetorical formats: the *two-part exposition* (containing a medial caesura) or the *continuous exposition* (lacking a successfully articulated medial caesura). In this chapter we are concerned with the former.

The Two-Part Exposition

This is the format most frequently employed by most composers of the second half of the eighteenth century. Hence when one confronts any sonata form from this period, the most reasonable initial expectation would be that one is about to encounter a two-part exposition. (As will be seen in chapter 4, a continuous exposition often plays upon, then overrides, this expectation.) The cardinal feature identifying this exposition type is the presence of a sufficiently deployed medial caesura and (often contrasting) second theme.

Part 1 comprises the establishment of the tonic and the energized drive to the medial caesura. It contains two action-spaces, the *primary-theme zone (P)* and the *transitional zone (TR)*, and culminates in the *medial caesura (MC)*, which we indicate by an apostrophe ('). Part 2 comprises the post-MC material and lasts until the end of the exposition. This section is concerned with the cadential affirmation of the new key (V in major-key sonatas, III or v in minor-key ones). Part 2 subdivides into the *secondary-theme zone (S)*—which normally concludes with the sounding of the first satisfactory perfect authentic cadence (PAC) in the new key that proceeds onward to differing material, the event that we